



ELSEVIER

Discrete Applied Mathematics 86 (1998) 3–25

DISCRETE
APPLIED
MATHEMATICS

Combinatorial variability of Vapnik–Chervonenkis classes with applications to sample compression schemes[☆]

Shai Ben-David^{*}, Ami Litman¹

Computer Science Department, Technion Haifa 32000, Israel

Received 6 January 1997; received in revised form 24 July 1997; accepted 23 October 1997

Abstract

We define embeddings between concept classes that are meant to reflect certain aspects of their combinatorial structure. Furthermore, we introduce a notion of *universal concept classes* – classes into which any member of a given family of classes can be embedded. These universal classes play a role similar to that played in computational complexity by languages that are hard for a given complexity class. We show that classes of half-spaces in \mathbb{R}^n are universal with respect to families of algebraically defined classes.

We present some combinatorial parameters along which the family of classes of a given VC-dimension can be grouped into sub-families. We use these parameters to investigate the existence of embeddings and the scope of universality of classes. We view the formulation of these parameters and the related questions that they raise as a significant component in this work.

A second theme in our work is the notion of *sample compression schemes*. Intuitively, a class C has a sample compression scheme if for any finite sample, labeled according to a member of C , there exists a short sub-sample so that the labels of the full sample can be reconstructed from this sub-sample.

By demonstrating the existence of certain compression schemes for the classes of half-spaces the existence of similar compression schemes for every class embeddable in half-spaces readily follows. We apply this approach to prove the existence of compression schemes for all ‘geometric concept classes’. © 1998 Elsevier Science B.V. All rights reserved.

0. Introduction

This paper introduces a partial order on set systems that comes to reflect certain aspects of their combinatorial richness. Much like the partial order of polynomial reducibility among computable languages, our combinatorial complexity partial order is based upon a notion of embeddings among set systems. Furthermore, we introduce a

[☆] This research was partially supported by the Technion fund for promotion of sponsored research.

^{*} Corresponding author. E-mail: shai@cs.technion.ac.il.

¹ E-mail: litman@cs.technion.ac.il.

notion of *universal set systems* that plays a role similar to that played in computational complexity by languages that are hard for a given complexity class.

Several parameters defined for a set system may be viewed as reflecting some aspects of the combinatorial richness of the system. Most notable among these parameters is the Vapnik–Chervonenkis dimension. Some other examples originating from computational learning theory are:

- The sample complexity of learning a concept class (in learning theory the term ‘concept class’ is used for a set system, we use both terms interchangeably).
- The size of the minimal compression scheme for a class. (We shall discuss several types of compression schemes, in the spirit of the schemes discussed by the Floyd and Warmuth [8], each of these schemes gives rise to its own parameter).
- The optimal mistake bound for learning the class online (sometimes called ‘the K-dimension’).

Each of these parameters induces a partial order over the family of concept classes. The partial orders we introduce here refine each of the partial orders induced by these parameters. In other words, if one class is embeddable in another (with respect to the embeddings discussed in this paper) then the value of each of the above parameters for the embedded class is at most the value this parameter has for the class it is embedded in.

Consequently, by demonstrating the existence of a certain compression scheme for a concept class which is universal for a given family of concept classes, the existence of similar compression schemes for all members of the family readily follows. We apply this approach to prove existence of compression schemes for all ‘geometric concept classes’ (a notion that will be defined precisely later).

Our investigation of the embedding relation among concept classes raises some basic questions concerning the structure of classes (or set systems) having a finite VC-dimension. We present some combinatorial parameters according to which the family of classes of a given VC-dimension can be grouped into sub-families. We believe that part of the contribution of this paper lies in the formulation of these parameters and in calling attention to the related questions, that seem to be both new and intriguing.

Embeddings of concept classes have been previously introduced in the context of computational learning theory. Pitt and Warmuth [14] introduce a notion of ‘prediction-preserving reductions’ for concept classes. Their aim is to define computationally efficient reductions among prediction problems. Consequently, their notion of reduction includes computational constraints. We shift the focus of the discussion from computational complexity issues to the combinatorial structure of classes. Our basic notion of embedding is similar to the Pitt–Warmuth reductions stripped off of their computational constraints. We introduce our notions of embeddings in Section 1 and in Section 2 we prove the universality of certain geometric classes and apply it to get some insight into the structure of algebraically defined classes.

Sample compression schemes were first introduced by Littlestone and Warmuth [12] and studied extensively by Floyd [7] and Floyd and Warmuth [8]. These papers establish close relationships between compression schemes and learning and provide sample

compression schemes for several families of concept classes. For geometric classes, [8] prove the existence of finite-size compression schemes for some sub-classes of the class of half-spaces in \mathbb{R}^n and [7] presents compression schemes for the classes of rectangles and triangles in \mathbb{R}^2 . In Section 3 we define several variants of sample compression schemes and discuss basic properties of such schemes. In Section 4, we introduce strong compression schemes for the (full) classes of half-spaces in \mathbb{R}^n , and finite-size compression schemes for any class of geometric objects in \mathbb{R}^n .

1. Embeddings of concept classes

1.1. Some preliminaries on classes and VC-dimension

Commonly, a concept class is defined as a collection of subsets of some domain set. For our purposes it will be instructive to define concept classes in a way that emphasizes the symmetry between ‘concepts’ and domain points.

Definition 1.

- A *concept class* is a triple $C = (X, Y, R)$ where X, Y are arbitrary sets and R is a binary relation, $R \subseteq X \times Y$, we call X the domain of C .
- A *concept* in such a class R is a subset of X of the form $c_y = \{x \in X : (x, y) \in R\}$, for some $y \in Y$.
- A collection C of subsets of X *shatters* a set $A \subseteq X$ if, for every $B \subseteq A$ there exists some $s \in C$ such that $s \cap A = B$.
- The *VC-dimension* of a class $C = (X, Y, R)$ is the supremum over all sizes of finite subsets of X shattered by $\{c_y : y \in Y\}$.
- The *dual class* of a concept class $C = (X, Y, R)$ is the class $C^D = (Y, X, R^D)$, where $R^D \subseteq Y \times X$ is defined by $(y, x) \in R^D \Leftrightarrow (x, y) \in R$. Note that a concept class C may be represented as a matrix over $\{0, 1\}$ whose (y, x) ’s entry is 1 iff $(x, y) \in R$ (to conform with common representations of concept classes as matrices, we chose to have rows corresponding to concepts—members of Y , and columns representing instances—members of X). If A_C is a matrix that represents a class C then its transposed matrix, A_C^T represents the dual class, C^D .

Definition 2.

For a pair of classes, $C = (X, Y, R)$ and $C' = (X', Y', R')$,

- We say that C' is a *subclass* of C if $X' \subseteq X, Y' \subseteq Y$ and R' is the restriction of R to $X' \times Y'$.
- Assuming $X = X'$ and $Y \cap Y' = \emptyset$, we define the *union* of the classes, $C \cup C' = (X, Y \cup Y', R \cup R')$.
- A class C is *finite* if both X and Y are finite sets.
- For any $A \subseteq X$, denote by $C|_A$ the subclass of C obtained by restricting C to the domain set A , i.e., $C|_A = (A, Y, R \cap (A \times Y))$.
- Let $F(C)$ denote the representation of C as a class of boolean functions on X :

$$F(C) = \{f_y : y \in Y\}$$

where, for any $y \in Y$,

$$f_y(x) = \begin{cases} 1 & \text{if } (x, y) \in R, \\ 0 & \text{otherwise.} \end{cases}$$

The basic combinatorial result concerning classes of finite VC-dimension is Sauer's lemma [15].

Lemma 1 (Sauer [15]). *If a class $C = (X, Y, R)$ has a finite VC-dimension, d , then for every finite $A \subseteq X$,*

$$|F(C|_A)| \leq \sum_{i=0}^d \binom{|A|}{i}$$

(We use the convention $\binom{m}{i} = 0$ for $i > m$).

We shall also need the following consequence of Sauer's lemma,

Claim 1. *For every pair of classes, $C_0 = (X, Y_0, R_0), C_1 = (X, Y_1, R_1)$ (over the same domain set X), $\text{VC-dim}(C_0 \cup C_1) \leq 2 \max\{\text{VC-dim}(C_0), \text{VC-dim}(C_1)\} + 1$.*

Proof. Let $k = \max(\text{VC-dim}(C_0), \text{VC-dim}(C_1))$, and assume, for contradiction, that $\text{VC-dim}(C_0 \cup C_1) > 2k + 1$. Let A be a subset of X of cardinality $2k + 2$ shattered by $C_0 \cup C_1$. This implies $|F((C_0 \cup C_1)|_A)| = 2^{2k+2} = \sum_{i=0}^{2k+2} \binom{2k+2}{i} = \binom{2k+2}{k+1} + 2 \sum_{i=0}^k \binom{2k+2}{i}$. Hence, for some $i \in \{0, 1\}$, $|F((C_i)|_A)| > \sum_{i=0}^k \binom{2k+2}{i}$. But, by Sauer's lemma, $|F((C_i)|_A)| \leq \sum_{i=0}^k \binom{2k+2}{i}$; a contradiction. \square

1.2. Embedding of classes

We turn now to the central tool of this paper – the notions of embeddings among concept classes. We define four variants of embeddings. Our basic definition (1 in the definition below) is similar to the definition of ‘prediction-preserving reductions’ of Pitt and Warmuth [14] stripped off of its computational constraints. As this paper considers combinatorial structure rather than computational complexity, these constraints are irrelevant to our discussion.

Definition 3. Given concept classes, $C = (X, Y, R)$ and $C' = (X', Y', R')$,

1. An *embedding* of C into C' is a pair of functions, $(\pi : X \mapsto X', \tau : Y \mapsto Y')$, such that for every $x \in X, y \in Y, (x, y) \in R$ iff $(\pi(x), \tau(y)) \in R'$.
2. A *generalized embedding* of C in C' is a pair of functions, $(\pi : X \mapsto X', \tau : Y \mapsto Y')$, and a function $\sigma : X \mapsto \{0, 1\}$ satisfying, for every $x \in X, y \in Y$,

$$\text{if } \sigma(x) = 0 \text{ then } (x, y) \in R \text{ iff } (\pi(x), \tau(y)) \in R',$$

and

if $\sigma(x) = 1$ then $(x, y) \in R$ iff $(\pi(x), \tau(y)) \notin R'$.

3. A class $C = (X, Y, R)$ is *weakly embeddable* in a class $C' = (X', Y', R')$ if every *finite* subclass of C is embeddable in C' . This notion applies to both of the above types of embeddings.

Note that all (four) notions of embeddability that arise from the above definition are reflexive and transitive relations (over a collection of concept classes). Consequently, they induce pre-orderings over concept classes, namely,

Notation 1.

1. $C \preceq_{\text{emb}} C'$ iff there exists an embedding of C into C' .
2. $C \preceq_{\text{gemb}} C'$ iff there exists a generalized embedding of C into C' .
3. $C \preceq_{\text{emb}}^* C'$ C is weakly embeddable in C' .
4. $C \preceq_{\text{gemb}}^* C'$ iff C is weakly embeddable in C' using generalized embeddings.

Note that $C \preceq_{\text{emb}} C'$, implies $C \preceq_{\text{gemb}}^* C'$, $C \preceq_{\text{gemb}} C'$ implies $C \preceq_{\text{emb}}^* C'$, and also $C \preceq_{\text{emb}} C'$ implies $C \preceq_{\text{gemb}} C'$ and $C \preceq_{\text{emb}}^* C'$ implies $C \preceq_{\text{gemb}}^* C'$.

Observation 1. Let F_d denote the class of all binary vectors over d entries. That is, $F_d = (X, Y, R)$ where $X = \{1, \dots, d\}$, $Y = \{0, 1\}^X$ and $(x, v) \in R$ iff $v(x) = 1$. The VC-dimension of F_d equals d , and for every class C of the same VC-dimension, $F_d \preceq_{\text{emb}} C$.

Many combinatorial parameters, reflecting aspects of the combinatorial richness of classes, are invariant under the above notions of embedding. The following claim lists some examples of such parameters that arise in the context of computational learning theory. We refer the reader to [4] for a definition of PAC-*learnability* and to [11] for a definition of *online learning*.

Claim 2.

1. $C \preceq_{\text{gemb}}^* C'$ implies $\text{VC-dim}(C) \leq \text{VC-dim}(C')$.
2. $C \preceq_{\text{gemb}}^* C'$ implies that, for any type of compression scheme discussed below, if C' has a compression scheme of some size, d , then so does C .
3. $C \preceq_{\text{gemb}}^* C'$ implies that, for every $\varepsilon, \delta > 0$ and $m \in \mathbb{N}$, if C' is (ε, δ) -PAC-learnable from m examples then so is C .
4. $C \preceq_{\text{gemb}}^* C'$ implies that, for every $m \in \mathbb{N}$, if C' is online learnable with at most m many mistakes, then so is C .

The proofs of the above claims follow quite easily from the relevant definitions. A proof of part 3 of the claim appears in [14].

The only statement in the claim that does not follow directly from the relevant definitions is part 2. The proof of part 2 involves a compactness consideration for compression schemes that will be discussed in Section 4.

Later on in the paper, we shall discuss some more parameters with the property that the partial orders they induce on classes are refined by \preceq_{gemb}^* . Consequently, one may view these ordering relations as reflecting the ‘combinatorial richness’ of classes.

Let us conclude this section by applying the notion of embedding to reprove a well known result relating the VC-dimension of a class to that of its dual class (we shall need this result later in Section 2).

Claim 3 (Laskowski, [10]). *For any class C :*

$$\log(\text{VC-dim}(C)) - 1 < \text{VC-dim}(C^D) < 2^{\text{VC-dim}(C)+1}.$$

Proof. As $(C^D)^D = C$ it suffices to prove only the first inequality. Let X be an arbitrary set of cardinality $k = \lfloor \log(\text{VC-dim}(C)) \rfloor$; let $Y \triangleq \{X' : X' \subset X\}$, $Z \triangleq \{Y' : Y' \subset Y\}$, $A \triangleq \langle X, Y, \in \rangle$ and $B \triangleq \langle Y, Z, \in \rangle$.

The class A satisfies $A \preceq_{\text{emb}} B^D$ via the mapping $\pi(x) \triangleq \{X' \subset X : x \in X'\}$ and $\tau(y) = y$. The class B is \preceq_{emb} -embeddable in any class of the same or greater VC-dim; hence, $B \preceq_{\text{emb}} C$, which implies $B^D \preceq_{\text{emb}} C^D$. Since \preceq_{emb} is transitive, $A \preceq_{\text{emb}} C^D$; hence $k = \text{VC-dim}(C^D)$. \square

Note that the inequalities of Claim 3 are tight, as demonstrated by the above class B .

2. Universal classes for families of geometric objects

The next natural concept to be defined is that of a class being an upper bound for a family of classes. This notion is analogous to that of a language being *hard* for a given complexity class in computational complexity theory.

Definition 4. Let C be a class, \mathcal{S} a family of classes and any partial order on classes, \preceq . We say that C is \preceq -*universal* for \mathcal{S} if, for every class $W \in \mathcal{S}$, $W \preceq C$.

In this section we shall present some families of classes and demonstrate the existence of universal classes for them with respect to the partial orders defined above.

2.1. Some concrete classes of geometric objects

We start our discussion of concrete classes by citing a definition that captures a wide family of ‘natural’ classes. This definition is due to Dudley [6] so we call members of this family *Dudley classes*.

Definition 5. Let \mathcal{F} be a collection of real-valued functions over some domain X . For a collection \mathcal{F} which is a vector space over the reals (with respect to pointwise addition and scalar multiplication) and any $h : X \mapsto \mathbb{R}$,

a *Dudley class* is a class of the form $D_{\mathcal{F},h} = (X, \mathcal{F}, R_{\mathcal{F},h})$, where

$$R_{\mathcal{F},h} = \{(x, f) : x \in X, f \in \mathcal{F}, \text{ and } f(x) + h(x) \geq 0\}.$$

In other words, letting $\text{pos}(g) \stackrel{\text{def}}{=} \{x \in X : g(x) \geq 0\}$, the concepts of $R_{\mathcal{F}, h}$ are exactly the sets $\{\text{pos}(f + h) : f \in \mathcal{F}\}$.

Dudley [6] proves that the VC-dimension of such a class, $R_{\mathcal{F}, h}$, equals the linear dimension of the vector space \mathcal{F} .

The collection of half-spaces in \mathbb{R}^n , the collection of all subsets of \mathbb{R} that are unions of at most k many intervals (for any fixed k), the collection of all subsets of \mathbb{R}^{n+1} that are defined by polynomial inequalities using fixed degree polynomials in $\{x_1, \dots, x_n\}$, are all examples of Dudley classes. The class of all n -dimensional balls, as well as many other natural classes of geometric objects, are subclasses of Dudley classes of the same VC-dimension. All the results obtained in this paper for Dudley classes apply to such classes as well.

We now introduce some concrete classes of geometric objects, these classes will be later shown to be universal w.r.t. families of Dudley classes.

Notation 2.

- Let HS^k denote the class of all half-spaces in \mathbb{R}^k , i.e.,

$$HS^k = (\mathbb{R}^k, \mathbb{R}^{k+1}, H) \text{ where } H = \{((x_1, \dots, x_k), (a_1, \dots, a_{k+1})) : \sum_{i=1}^k a_i x_i + a_{k+1} \geq 0\}$$

- Let PHS^k denote the class of *positive* half-spaces in \mathbb{R}^k . That is, we restrict the class HS^k by replacing \mathbb{R}^{k+1} by $\{(a_1, \dots, a_{k+1}) : a_1 > 0\}$.
- Let PHS_0^k denote the class of positive half-spaces in \mathbb{R}^k that pass through the origin (the zero vector of \mathbb{R}^k). That is, we restrict the set of concepts further by allowing only vectors $(a_1, \dots, a_k, a_{k+1})$ in which the last component, a_{k+1} , is zero.

It should be noted that these classes are all Dudley classes. The following is well known [7].

Claim 4. For every $k \in \mathbb{N}$,

1. $\text{VC-dim}(HS^k) = k + 1$.
2. $\text{VC-dim}(PHS^k) = k$.
3. $\text{VC-dim}((HS^k)^D) = \text{VC-dim}((PHS^k)^D) = k$.
4. $\text{VC-dim}(PHS_0^k) = k - 1$.

2.2. Universality theorems

Theorem 1. Let k be any natural number and let \mathcal{D}_k denote the family of Dudley classes of dimension k .

1. PHS_0^{k+1} is \preceq_{emb} -universal for \mathcal{D}_k .
2. $(HS^k)^D$ is \preceq_{emb} -universal for \mathcal{D}_k .
3. PHS^k is \preceq_{gemb}^* -universal for \mathcal{D}_k .
4. $(PHS^k)^D$ is \preceq_{gemb} -universal for \mathcal{D}_k .

Proof.

1. Let $D_{\mathcal{F},h} = (X, \mathcal{F}, R)$ be a k -dimensional Dudley class and let (f_1, \dots, f_k) be a linear basis for the space \mathcal{F} . Now, for $f = \sum_{i=1}^k \alpha_i f_i$, let

$$\tau(f) = (\alpha_1, \dots, \alpha_k, 1, 0)$$

and, for any $x \in X$ let

$$\pi(x) = (f_1(x), \dots, f_k(x), h(x)).$$

2. (Using the notation of the previous item) For $f \in \mathcal{F}$, let

$$\tau(f) = (\alpha_1, \dots, \alpha_k)$$

and, for $x \in X$, let

$$\pi(x) = (h(x), f_1(x), \dots, f_k(x)).$$

3. Applying part 1 of this theorem, it suffices to prove that $PHS_0^{k+1} \preceq_{\text{gemb}}^* PHS^k$. The proof of this statement is rather technical and is deferred to the Appendix for the sake of readability of this section.
4. Applying part 2 of this theorem, it suffices to prove that $(HS^k)^D \preceq_{\text{gemb}}^* (PHS^k)^D$. Let (A, B, H) be a finite subclass of $(HS^k)^D$, due to the finiteness of the class, we may replace each half-space $a \in A$ by a half-space a' that induces the same partition over the points of B and so that no point of B is on its boundary hyper-plane. Now let τ be the identity function on B , for each $a \in A$, let $\pi(a)$ be the positive half-space defined by the boundary of a' , and let $\sigma(a)$ be 0 if a is a positive half-space and 1 otherwise. \square

A natural question that arises at this stage is if the universality of PHS^k and $(PHS^k)^D$ holds with respect to wider classes. By Claim 2, we know that no class whose VC-dimension exceeds k can be embedded in a class having VC-dimension k , but is it the case that PHS^k and $(PHS^k)^D$ are universal for the family of all classes of VC-dimension k ? Alon, Frankl and Rodel [1] use a counting argument to show that there exist, for every n , concept classes of size $n \times n$ (i.e., both the domain set X and the set of concepts Y have cardinality n) that are not \preceq_{emb} -embeddable in PHS^k unless $k \geq n/32$. As the VC-dimension of such a class cannot exceed $\log(n)$, it follows that some classes cannot be embedded in PHS^k unless k is exponentially large than their VC-dimension. Ben-David et al. [3] have recently extended these results to prove that for every $d > 10$ and every $k \in \mathbb{N}$ there exists a finite class C_d^k having VC-dimension d and yet being not embeddable in \mathbb{R}^k . These counting arguments are strong enough to guarantee non-generalized-embeddability. Note also that, as the classes C_d^k are all finite, their nonembeddability is equivalent to being non-weakly-embeddable.

The combinatorial parameters discussed in the following subsection will provide several ways to show non-embeddability in HS^k for some concrete and rather simple classes.

2.3. Implications to VC-dimension related parameters

The above results imply that, for many combinatorial parameters, the values of these parameters for the class of positive half-spaces (and its dual class) are upper bounds on the value of these parameters for any Dudley class of the same dimension. In Section 4 below we shall discuss sample compression schemes for such classes. Let us mention here some other parameters for which these considerations yield new results.

The following notion is discussed in [7] and is attributed there to an unpublished manuscript of E. Welzl.

Definition 6. A concept class $C = (X, Y, R)$ having some finite VC-dimension, k , is a *maximum class* if for every finite $A \subseteq X$ the cardinality of $F(C|_A)$ meets the upper bound of Sauer's lemma. (That is, for $|A| \leq k$, $|F(C|_A)| = 2^{|A|}$, and for $|A| > k$, $|F(C|_A)| = \sum_{i=0}^k \binom{|A|}{i}$).

Floyd and Warmuth [7, 8] have shown that for any set $X \subset \mathbb{R}^k$ whose points are in general position,² $(PHS^k)|_X$ is a maximum class. Just the same, there are Dudley classes which do not have any nontrivial maximum subclass, namely:

Claim 5. Every subclass of HS^k with VC-dimension $k + 1$, whose domain set has cardinality greater than $(k + 1)^2 + k$, is not a maximum class.

Proof. Clearly, over any finite subset of \mathbb{R}^k , there are at most twice as many half-spaces than positive half-spaces. Let C be a subclass of HS^k satisfying the claims assumptions. Let A be any finite subset of the domain of C of cardinality $> (k + 1)^2 + k$. As $\text{VC-dim}(PHS^k) = k$, Sauer's Lemma implies that $|F(PHS^k|_A)| \leq \sum_{i=0}^k \binom{|A|}{i}$. It follows that $|F(C|_A)| \leq 2 \sum_{i=0}^k \binom{|A|}{i} < \sum_{i=0}^{k+1} \binom{|A|}{i}$. \square

On the other hand, we have:

Claim 6. Every Dudley class of dimension k is \preceq_{gemb}^* embeddable in a maximum class of VC-dimension k .

Proof. Let X be a dense subset of \mathbb{R}^k in general position.³ By [7], $(PHS^k)|_X$ is a maximum class. Let D be a Dudley class of dimension k . By Theorem 1, D is \preceq_{gemb}^* embeddable in PHS^k . Let D' be a finite subclass of D . Clearly, D' is embeddable in

² General position means that, for every $i \leq k$, any i -dimensional hyper-plane contains at most i points of X .

³ The existence of such a set follows from a standard infinite counting argument; Let $\{B_x : x < 2^{\aleph_0}\}$ enumerate all open balls in \mathbb{R}^k . Define the sequence of points of X by (infinite) induction. Assuming $\{x_\beta : \beta < \alpha\}$ is already defined, pick $x_\alpha \in B_\alpha \setminus \{x_\beta : \beta < \alpha\}$, so that x_α does not share a hyper-plane with any $(k - 1)$ -tuple of point from $\{x_\beta : \beta < \alpha\}$. This is always possible as the collection of forbidden hyper-planes has cardinality at most $|\alpha| \times \aleph_0$, which is strictly less than 2^{\aleph_0} , and no non-empty ball in \mathbb{R}^k can be covered by a collection of hyper-planes of cardinality less than continuum.

a finite subclass of PHS^k . Since X is dense, D' is embeddable in a finite subclass of $(PHS^k)_{|X}$ as well. \square

The above claim does not hold for general (not Dudley) classes. Floyd and Warmuth [8] have presented examples of finite classes that cannot be extended to maximum classes of the same VC-dimension. In our terminology it means that they are not \preceq_{emb} -embeddable⁴ in any maximum class of the same VC-dimension.

Towards gaining a better understanding of the embeddability partial orders, we introduce two more parameters of a class C . These parameters are the VC-dimensions of images of C under two simple mappings of classes – mapping a class C to its dual class C^D and mapping a class to its completion \bar{C} , defined as follows.

Notation 3. Given a class $C = (X, Y, R)$, let \bar{R} be the relation over $X \times (Y \times \{0, 1\})$ defined by

$$(x, (y, 0)) \in \bar{R} \text{ iff } (x, y) \in R, \text{ and } (x, (y, 1)) \in \bar{R} \text{ iff } (x, y) \notin R$$

We define the *completion* of a class C as the class $\bar{C} = (X, (Y \times \{0, 1\}), \bar{R})$.

Claim 7. For any class, $\text{VC-dim}(\bar{C}) \leq 2(\text{VC-dim}(C)) + 1$.

Proof. Note that \bar{C} is the union of $C_0 = (X, Y \times 0, \bar{R})$ and $C_1 = (X, Y \times 1, \bar{R})$. Since $\text{VC-dim}(C) = \text{VC-dim}(C_0) = \text{VC-dim}(C_1)$, by Claim 1: $\text{VC-dim}(\bar{C}) \leq 2(\text{VC-dim}(C)) + 1$. \square

The following claim extends Claim 2 to the new parameters. Namely, it relates the embeddings partial orders to those induced by the VC-dimension of the dual of a class and of its completion.

Claim 8.

1. $C_1 \preceq_{\text{gemb}}^* C_2$ implies $\bar{C}_1 \preceq_{\text{gemb}}^* \bar{C}_2$, and therefore, $\text{VC-dim}(\bar{C}_1) \leq \text{VC-dim}(\bar{C}_2)$.
2. $C_1 \preceq_{\text{emb}}^* C_2$ implies $C_1^D \preceq_{\text{emb}}^* C_2^D$, and therefore, $\text{VC-dim}(C_1^D) \leq \text{VC-dim}(C_2^D)$.
3. $C_1 \preceq_{\text{gemb}}^* C_2$ implies $\bar{C}_1^D \preceq_{\text{emb}}^* \bar{C}_2^D$, and therefore, $\text{VC-dim}(C_1^D) \leq \text{VC-dim}(C_2^D)$.

Corollary 2. For every finite k ,

1. If C is a k -dimensional Dudley class then $\text{VC-dim}(\bar{C}) \leq k + 1$.
2. If C is a k -dimensional Dudley class then $\text{VC-dim}(C^D) \leq k + 1$.

Proof.

1. Applying part 3 of Theorem 1, we know that $C \preceq_{\text{gemb}}^* PHS^k$. The claim now follows by applying part 1 of Claim 8 and noting that $\overline{PHS^k} \preceq_{\text{emb}}^* HS^k$.⁵

⁴ Note that, for finite classes, embeddability and weak-embeddability coincide.

⁵ The reason we write ' \preceq_{emb}^* ' rather than ' \preceq ' is only formalistic – we have defined HS^k and PHS^k as classes of closed half-spaces, so the complement of an element of PHS^k is not a member of HS^k . In Section 4 we shall see that this difference can be neglected for our purposes.

2. Applying part 2 of Theorem 1, we have $C \preceq_{\text{emb}}^* (HS^k)^D$, by part 2 of Claim 8 we conclude that $\text{VC-dim}(C^D) \leq \text{VC-dim}(((HS^k)^D)^D)$. Now just note that, for every class, $((C)^D)^D = C$ and recall that $\text{VC-dim}(HS^k) = k + 1$. \square

In the following examples, we apply Claim 8 to obtain some non-embeddability results for natural classes.

Example 1. Recall that F_d denotes the class of all binary vectors over d entries ($F_d = (X, Y, R)$ where $X = \{1, \dots, d\}$, $Y = \{0, 1\}^X$ and $(x, v) \in R$ iff $v(x) = 1$) $\text{VC-dim}(F_d) = d$ while $\text{VC-dim}((F_d)^D) = \lfloor \log(d) \rfloor$ (as $|Y| = d$ it cannot exceed $\lfloor \log(d) \rfloor$, and as its dual class has VC-dimension d , Claim 3 implies that it cannot be smaller than $\lfloor \log(d) \rfloor$). It follows that $(F_d)^D$ is not \preceq_{gemb}^* -embeddable in PHS^k for any k less than d . Consequently, $(F_{2^d})^D$ is an example of a class whose VC-dimension is d and it is not embeddable in any Dudley class of dimension $< 2^d - 1$.

Example 2. For any $k \leq n \in \mathbb{N}$, let HW_k^n denote the class of binary vectors of length n having at most k many non-zero entries (i.e., the class whose representing matrix, $A_{HW_k^n}$, has these vectors as its rows).

Note the $\text{VC-dim}(HW_k^n) = k$ and, whenever $n \geq 2k + 1$, $\text{VC-dim}(\overline{HW_k^n}) = 2k + 1$. It follows that HW_k^n is not \preceq_{gemb}^* -embeddable in PHS^d for any $d \leq 2k - 1$.

We therefore have, for every value of k , an example of a rather simple class of VC-dimension k that is not embeddable in PHS^{2k-1} and, consequently, cannot be extended to any d -dimensional Dudley class, for any $d < 2k$.

On the other hand, the class HW_k^n is \preceq_{emb} -embeddable in a $2k$ -dimensional Dudley class as follows. Let I^k denote the class of the union of at most k line intervals, i.e., $I^k \triangleq (\mathbb{R}, Y, \in)$ where $Y = \{A \subset \mathbb{R} : A \text{ is the union of at most } k \text{ closed intervals}\}$. Clearly, for any n , $HW_k^n \preceq_{\text{emb}} I^k$. The class I^k is a $2k$ -dimensional Dudley class, $D_{\mathcal{F}, h}$ where $h(x) = x^{2k}$ and \mathcal{F} the vector space of polynomials whose degree is at most $2k - 1$.

3. Sample compression schemes

Given a class of sets, $C = (X, Y, R)$, a sample-compression scheme for C is a mapping from (input) sequences of elements of X ('examples'), labeled according to the truth value of $R(x, y)$ for some fixed $y \in Y$, to short sub-sequences. It is required that the labels of the examples in the input sequence can be reconstructed from the short sub-sequence. The *size* of a sample compression scheme for a class C is the (minimal) upper bound on the length of the sub-sequences in the range of this mapping.

Sample compression schemes were first introduced by Littlestone and Warmuth [12] and studied extensively by Floyd [7] and Floyd and Warmuth [8]. These papers establish close relationships between compression schemes and learning and provide sample compression schemes for several families of concept classes.

A simple counting arguments shows that, up to a constant factor, the VC-dimension of a class C is a lower bound on the size of compressed sub-sequences of any

compression scheme for C . The question whether the VC-dimension determines also an upper bound (on the minimal sufficient size of compressed images) is an intriguing open problem.

In this section we shall show that, for all Dudley classes, compression schemes of size equal to their VC-dimension do exist. Furthermore, for all geometric classes we can find finite-size compression schemes.

We shall discuss three variants of this notion, the first two – *labeled* compression and *unlabeled* compression – were introduced in the works of Floyd, Littlestone and Warmuth mentioned above, while the third – *array* compression – is a new variant of the previous notions. For the clarity of representation, let us begin by formally defining the basic version of sample compression schemes.

Notation 4.

- For partial functions f, g , let $g \sqsubseteq f$ denote that f is an extension of g (i.e. $\text{Dom}(g) \subseteq \text{Dom}(f)$ and, for $x \in \text{Dom}(g)$, $g(x) = f(x)$).
- Recall the $F(C)$ denotes the representation of a class C as a class of boolean functions over its domain X (Definition 2). Let $F(C)^{<\infty}$ denote the set of finite partial functions of functions in $F(C)$,

$$F(C)^{<\infty} = \{g : \text{Dom}(g) \text{ is finite and, for some } f \in F(C), g \sqsubseteq f\}$$

Finally,

- For any finite d , let $F(C)^{\leq d}$ denote the set of functions in $F(C)^{<\infty}$ whose domain is of size $\leq d$.

Definition 7 (Floyd [7]). A *size- d labeled compression scheme* for a class C is a mapping

$$H : F(C)^{\leq d} \mapsto \{0, 1\}^X$$

such that, for any $f \in F(C)^{<\infty}$ there exists some $g \in F(C)^{\leq d}$, so that

$$q \sqsubseteq f \sqsubseteq H(g)$$

A function H as in the above definition is called a *reconstruction function*. We have chosen to define compression schemes via their reconstruction functions. Clearly, given a reconstruction function H as above, a converse compression function (mapping f to an appropriate g) exists as well.

Let us show a few simple examples:

Example 3. For $k \leq n \in \mathbb{N}$, recall that HW_k^n denotes the class of binary vectors of length n whose Hamming weight is at most k . The function H defined by

$$H(u)(i) = \begin{cases} 1 & \text{if } u(i) = 1 \\ 0 & \text{otherwise} \end{cases}$$

is a size- k compression scheme for HW_k^n .

It turns out that the above compression scheme is a special case of a scheme that applies to a wide family of classes, namely,

Example 4. Let $C = (X, Y, R)$ be an *intersection-closed* class, i.e., for every $y_1, y_2 \in Y$ there exists $y_3 \in Y$ such that $c_{y_3} = c_{y_1} \cap c_{y_2}$. Define a reconstruction function, H , by

$$H(g)(x) = 1 \text{ iff } x \in \bigcap \{c_y : g^{-1}(1) \subseteq c_y\}$$

Clearly, for all $g \in F(C)^{<\infty}$, $g \sqsubseteq H(g)$. Natarajan [13] proves that for intersection closed classees C ,

$$\forall f \in F(C)^{<\infty}, \exists A \subseteq f^{-1}(1), \text{ so that } |A| \leq \text{VC} - \dim(C)$$

$$\text{and } f^{-1}(1) \subseteq \bigcap \{c_y : A \subseteq c_y\}$$

It follows that H is a compression scheme of size equal to the VC-dimension of C (provided C is intersection-closed).

Note that in both of the examples above the compression scheme, H , is actually stronger than what is required by Definition 7. The function H in these examples can be restricted to positive examples only – it uses only a subsample of f 's domain on which $f(x) = 1$. Examples as these motivate the following definition of a stronger type of compression schemes.

For a set X , let $\mathcal{P}(X)$ denote the power set of X (that is, the set of all subsets of X), and let $\mathcal{P}(X)^{\leq d}$ denote the set of all subsets of X whose cardinality is at most d .

Definition 8 (Floyd [8]). A *size- d unlabeled compression scheme* for a class C is a mapping

$$H : \mathcal{P}(X)^{\leq d} \mapsto \{0, 1\}^X$$

such that, for any $f \in F(C)^{<\infty}$ there exists some $A \in \mathcal{P}(X)^{\leq d}$, such that

$$A \subseteq \text{Dom}(f) \text{ and } f \sqsubseteq H(A)$$

Note that the compression schemes in Examples 3 and 4 above are, in fact, unlabeled compression schemes.

In this work we shall discuss yet another variant of compression schemes.

Definition 9. Given a concept class $C = (X, Y, R)$, let \mathbf{b} be any symbol such that $\mathbf{b} \notin X \cup Y$. A *size- d array compression scheme* for a class C is a mapping

$$H : (X \cup \{\mathbf{b}\})^d \mapsto \{0, 1\}^X$$

such that, for any $f \in F(C)^{<\infty}$ there exists some $\sigma \in (X \cup \{\mathbf{b}\})^d$,

$$\text{Range}(\sigma) \subseteq \text{Dom}(f) \cup \{\mathbf{b}\} \text{ and } f \sqsubseteq H(\sigma)$$

(where, for a sequence $\sigma = (1, \dots, a_d)$, $\text{Range}(\sigma)$ is the set $\{1, \dots, a_d\}$).

Array compression may be viewed as reflecting a computer program that stores the information gathered by viewing a long sequence of labeled examples (the function f) in a d -size array memory in which every cell may either contain one of the examples seen in f or may be left blank. Note that, as the domain of H in array compression schemes is a set of *sequences* (rather than just sets), such schemes can encode information into the ordering of the stored examples.

The following lemmata are straightforward.

Lemma 2. *If a class C has an unlabeled-compression scheme of size k then it has an array-compression scheme of size k .*

Lemma 3. *If a class C has a labeled-compression of size k then it has an array-compression scheme of size $2k$.*

Array compression schemes may be viewed as a special case of a more general definition of compression. Floyd and Warmuth introduce *extended* sample compression schemes that are allowed to store, on top of a finite set of sample points from the domain of the compressed function (f), a finite set of bits.

It follows that any result concerning these extended schemes is valid for array compression schemes as well. In particular, this applies to the results of Littlestone and Warmuth [12] that the existence of compression schemes of finite size implies PAC learnability, and to the results of Ben-David, Bshouty and Kushilevitz [2], showing that the existence of computationally efficient compression schemes⁶ for a class C implies the existence of efficient Online Learning algorithms for the class of all boolean combinations of members of C .

Floyd [7] and Floyd and Warmuth [8] establish the existence of labeled compression schemes for all maximum classes to size equal to their VC-dimension. This implies, as a special case, sample compression of size k to the class of positive half-spaces over a set of points in general position in \mathbb{R}^k . Floyd [7] also presents size 5 and 6 labeled-sample compression schemes for the class of rectangles in \mathbb{R}^2 and for the class of triangles in \mathbb{R}^2 (respectively). In the next section we shall introduce, for any $k \in \mathbb{N}$, $k + 1$ -size *un-labeled* compression schemes for any class of geometric objects in \mathbb{R}^k .

Before we do that, we develop in the next two subsections, some basic technical tools concerning compression schemes.

3.1. A compactness lemma for compression schemes

Lemma 4. *Let C be a class, n an integer and τ one of the following types of compression schemes: ‘unlabeled’, ‘labeled’ or ‘array’. Then if any finite subclass of C admits a type τ compression scheme of size n then so does C .*

⁶ A compression scheme is *computationally efficient* if the function H from samples to (representations of) boolean functions over X can be computed in polynomial time (in the input sample size). We do not discuss the computational complexity aspects of compression schemes in this paper.

Proof. We consider only the case of array compression. The proof for the other cases is similar. Our proof is based on the Compactness Theorem of Predicate Logic.

Let $C = \langle X, Y, R \rangle$ be a class s.t. any finite subclass of C admits an array compression of size n . Assume, w.l.o.g., that $X \cap Y = \emptyset$ and $\mathbf{b} \notin X \cup Y$. Let $Z = X \cup \{\mathbf{b}\}$.

Consider the two-sorted model $\mathfrak{U} = \langle Z, Y; X, R, \mathbf{b}, x, y \rangle_{x \in X, y \in Y}$. This is, the universe of \mathfrak{U} is $Z \cup Y$; \mathfrak{U} has the unary predicates Z, Y and X , the binary predicate R , the constant \mathbf{b} , the constant x for each $x \in X$ and the constant y for each $y \in Y$. The language of \mathfrak{U} has the unary predicated \hat{Z}, \hat{Y} and \hat{X} , the binary predicates ‘=’ and \hat{R} , the constant symbol $\hat{\mathbf{b}}$, and constant symbols \hat{x} and \hat{y} for any $x \in X$ and $y \in Y$.

Let \hat{H} be an $(n+1)$ -ary predicate symbol. For an integer k , let ψ_k be the sentence:

$$\begin{aligned} \psi_k \triangleq & \forall y \in \hat{Y} \forall x_1, \dots, x_k \in \hat{X} \exists z_1, \dots, z_n \in \hat{Z} : \left[\bigwedge_{i=1}^n \bigvee_{j=1}^k (z_i = \hat{\mathbf{b}} \vee z_i = x_j) \right] \\ & \times \bigwedge_{j=1}^k \left[\hat{H}(z_1, \dots, z_n, x_j) \iff \hat{R}(x_j, y) \right]. \end{aligned}$$

Define $\Gamma = \{\psi_k : k = 1, 2, \dots\}$. Note that Γ “says” that the class admits an array compression of size n via the function $\bar{H} : z^n \mapsto \{0, 1\}^X$ defined by

$$\bar{H}(\bar{z})(x) = 1 \iff H(\bar{z}, x).$$

Let $\Delta_{\mathfrak{U}}$ be the diagram of \mathfrak{U} – the set of all atomic sentences and negation of atomic sentences which hold in \mathfrak{U} . Define $\Gamma^+ = \Gamma \cup \Delta_{\mathfrak{U}}$. Since any finite subclass of C has an array compression of size n , it follows that any finite subset of Γ^+ has a model. By the Compactness Theorem, Γ^+ has a model.

Let $\mathfrak{U}' = \langle Z', Y', X', R', \mathbf{b}', x', y', H' \rangle_{x \in X, y \in Y}$ be a model of Γ^+ . Since $\Delta_{\mathfrak{U}} \subset \Gamma^+$, we may assume that \mathfrak{U} is a sub-model of \mathfrak{U}' . Consider the class $C' = \langle X', Y', R' \rangle$. The class C' admits an array compression of size n , and C is a subclass of C' . Hence, C admits such a compression. \square

A consequence of Lemma 4 is:

Lemma 5. *Let τ be one of the following types of compression schemes: ‘unlabeled’, ‘labeled’ or ‘array’. Assume C admits a τ -compression of size n and $C' \preceq_{\text{gemb}}^* C$. Then C' admits a τ -compression of size n .*

3.2. Cartesian product of classes

In this section we define the Cartesian Product of classes and show that if all the component classes admit an array compression of a finite size, then so does their product. (It is yet unknown whether the same holds for any other type of compression.) In Section 5 below we use this result to establish that all the geometric classes admit an array compression.

Let us henceforth assume that the logical value *True* is 1 and *False* is 0. We use the following notation to denote sequences and their elements. \bar{u} denotes a sequence of elements; u_i or $(u)_i$ denotes the i -th elements of \bar{u} ; $\bar{\bar{v}}$ denotes a sequence of sequences; \bar{v}_i or $(\bar{v})_i$ denotes the i -th sequences of $\bar{\bar{v}}$, and $v_{i,j}$ or $(v_i)_j$ denotes the j -th element of the i -th sequence of $\bar{\bar{v}}$.

Definition 10. Let $C_1 = \langle X_1, Y_1, R_1 \rangle, \dots, C_k = \langle X_k, Y_k, R_k \rangle$ be classes and f a k -ary Boolean function (i.e., $f : \{0, 1\}^k \mapsto \{0, 1\}$). The f -Cartesian product of C_1, \dots, C_k , denoted $(C_1 \times C_2 \times \dots \times C_k)^f$ is the class $C = \langle X, Y, R \rangle$ where

$$X \triangleq X_1 \times X_2 \times \dots \times X_k, Y \triangleq Y_1 \times Y_2 \times \dots \times Y_k$$

and

$$R(\bar{x}, \bar{y}) = f(R_1(x_1, y_1), R_2(x_2, y_2), \dots, R_k(x_k, y_k))$$

Example 5. Let f be the ‘and’ function and let C be the class of line intervals, i.e., $C = \langle \mathbb{R} \{[x, y] : x, y \in \mathbb{R} \wedge x < y\}, \in \rangle$. Then $(C \times C)^f$ is the class of axes parallel rectangles.

Lemma 6. Let C_1, \dots, C_k and C'_1, \dots, C'_k be classes s.t. $C_i \preceq_{\text{emb}} C'_i$ for all $1 \leq i \leq k$, and let f be a k -ary Boolean function. Then $(C_1 \times \dots \times C_k)^f \preceq_{\text{emb}} (C'_1 \times \dots \times C'_k)^f$.

The same holds for \preceq_{emb}^* instead of \preceq_{emb} .

Proof. The proof is immediate. \square

The above lemma does not hold for \preceq_{gemb} or \preceq_{gemb}^* . As a counter example, consider the classes C and C' represented by the following matrices, where rows represent concepts and columns represent domain points:

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad C' = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Clearly, $C \preceq_{\text{gemb}}^* C'$. Let f be the ‘and’ function. Then

$$(C \times C)^f = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (C' \times C')^f = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Clearly, $(C \times C)^f$ is not \preceq_{gemb} than $(C' \times C')^f$.

We next show that if all the components of a Cartesian product admit an array compression, then so does the product.

Lemma 7. Let C_1, \dots, C_k be classes s.t. each C_i admits an array compression of size n_i , and let f be a k -ary Boolean function. Then $(C_1, \dots, C_k)^f$ admits an array compression of size $\sum_{i=1}^k n_i$.

Proof. For $i = 1, \dots, k$, let $C_i = \langle X_i, Y_i, R_i \rangle$ and let $H_i : (X_i \cup \{\mathbf{b}\})^{n_i} \mapsto \{0, 1\}^X$ be a reconstruction function of C_i . Let $(C_1 \times \dots \times C_k)^f = C = \langle X, Y, R, \rangle$, $n = \sum_{i=1}^k n_i$, and define $m_j = \sum_{i=1}^j n_i$.

For $\bar{z} \in (X \cup \{\mathbf{b}\})^n$, $1 \leq i \leq k$, and $1 \leq j \leq l \leq n$, define: $\bar{z}[j, l](i) \triangleq \langle z_{j,i}, z_{j+1,i}, \dots, z_{l,i} \rangle$ where, in this definition, $\mathbf{b}_i = \mathbf{b}$.

Define $H : (X \cup \{\mathbf{b}\})^n \mapsto \{0, 1\}^X$ by:

$$H(\bar{z})(\bar{x}) = f \left(H_1(\bar{z}[1, m_1](1))(x_1), H_2(\bar{z}[m_1 + 1, m_2](2))(x_2), \dots, H_k(\bar{z}[m_{k-1} + 1, m_k](k))(x_k) \right).$$

To show that H is a reconstruction function for C , let X' be a finite subset of X and $\bar{y} \in Y$.

For $1 \leq i \leq k$, define $X'_i \triangleq \{x_i : \bar{x} \in X'\}$. Since X'_i is a finite subset of X_i and H_i a reconstruction function for C_i , there is $\bar{u}^i \in (X'_i \cup \{\mathbf{b}\})^{n_i}$ s.t. $H_i(\bar{u}^i)(x) = R_i(x, y_i)$ for any $x \in X'_i$. There is $\bar{z} \in (X' \cup \{\mathbf{b}\})^n$ s.t.: $\bar{z}[m_{i-1} + 1, m_i](i) = \bar{u}^i$ for all $1 \leq i \leq k$. For this \bar{z} and any $\bar{x} \in X'$:

$$\begin{aligned} H(\bar{z})(\bar{x}) &= f \left(H_1(\bar{z}[1, m_1](1))(x_1), H_2(\bar{z}[m_1 + 1, m_2](2))(x_2), \dots, \right. \\ &\quad \left. H_k(\bar{z}[m_{k-1} + 1, m_k](k))(x_k) \right) \\ &= f \left(H_1(\bar{u}^1)(x_1), H_2(\bar{u}^2)(x_2), \dots, H_k(\bar{u}^k)(x_k) \right) \\ &= f(R_1(x_1, y_1), R_2(x_2, y_2), \dots, R_k(x_k, y_k)) \\ &= R(\bar{x}, \bar{y}). \quad \square \end{aligned}$$

4. Compressing the dual class of half-spaces

In this section we show that $(HS^n)^D$, the dual class of half-spaces of \mathbb{R}^n , admits an unlabeled compression scheme of size n . This implies that any Dudley class C admits an unlabeled compression scheme of size $\text{VC-dim}(C)$.

Let us begin by an intuitive overview of our compression scheme, we shall then go on to a precise formal presentation. Note that a sample in $(HS^n)^D$ is a finite collection of half-spaces labeled according to the way they classify the target point. Such a sample forms a ‘cell’ in \mathbb{R}^n , (and the target point is in this cell). Any point inside this cell gives rise to the same classification of all the sample’s half-spaces.

Applying the compactness lemma (Lemma 4), we may settle for designing a compression scheme for finite classes only. For such a classes, we show that one may assume that the points and half-spaces of the class are in ‘general position’. Given a class that meets these assumptions, we show that there always exists some reference point t that has some desired property (Definition 14 below) with respect to our class. Once such a point is fixed our compression scheme works as follows: given a sample let g be the cell it defines and let $m_t(g)$ be the point on the cells boundary which is

closest to t . We show that such a point is determined by at most n many half-spaces for the sample, and use $m_t(g)$ to determine labels to half-spaces (according to the way they classify this point). As the target point and $m_t(g)$ are in the same cell defined by the sample, such a labeling scheme agrees with the original labels on each of the sample's half-spaces.

Let us turn now to a rigorous presentation of the compression scheme. We consider \mathbb{R}^n to be an n -dimensional linear space over the real field and a metric space, where the metric is the Euclidean distance d . The dimension n is fixed throughout this section.

Definition 11. We refer to members of \mathbb{R}^n as *points*. A *j -dimensional coset* is a nonempty subset of \mathbb{R}^n of the form $y + q$ where $y \in \mathbb{R}^n$ and q is a j -dimensional linear subspace of \mathbb{R}^n . A singleton is a 0-dimensional coset, a *line* is a 1-dimensional coset, and a *hyper-plane* is an $(n - 1)$ -dimensional coset. A *half-space* is a subset of \mathbb{R}^n of the form

$$\{\alpha \in \mathbb{R} : \alpha \geq \beta\} \cdot y + q,$$

where q is an $(n - 1)$ -dimensional linear subspace, y is a point not in q , and $\beta \in \mathbb{R}$. So, a half-space is a closed convex subset of \mathbb{R}^n whose boundary is a hyper-plane. For any hyper-plane there are exactly two half-spaces whose boundaries are that hyper-plane; we refer to two such half-spaces as *adjacent* half-spaces. Let H^n denote the set of all half-spaces of \mathbb{R}^n .

In this section we use the following definition of $(HS^n)^D$.

$$(HS^n)^D \triangleq \langle X, Y, R \rangle; X \triangleq H^n \cup \{\mathbb{R}^n, \emptyset\}, \quad Y \triangleq \mathbb{R}^n \text{ and } R(x, y) \triangleq y \in x.$$

This $(HS^n)^D$ differs from its namesake of Notation 2; however, the two classes are \preceq_{emb} -embeddable one in the other.

We show that $(HS^n)^D$ admits an unlabeled compression scheme⁷ of size n . To this end, we build a small machinery which is based on linear algebra, and use the following elementary lemmas whose proof is omitted.

Lemma 8. Let q and p be cosets of dimensions j and $n - 1$. Then $q \cap p$ is either empty or a coset of dimension j or $j - 1$.

For $q \subset \mathbb{R}^n$ let $B(q)$ denote the boundary of q . For Q , a set of subsets of \mathbb{R}^n , define $B(Q) \triangleq \{B(q) : q \in Q\}$.

Lemma 9. Let L be a finite set of lines, g a finite set of points and s a half-space. Then there is a half-space s' s.t.

1. $B(s') \cap l$ is a singleton for any $l \in L$.

⁷ As far as compression is concerned, it is clearly irrelevant whether $(HS^n)^D$ is composed of the open half-spaces the closed half-spaces or both.

2. $s' \cap g = s \cap g$.
3. $B(s') \cap g = \emptyset$.

Definition 12. Let t be a point and g a closed, convex and nonempty subset of \mathbb{R}^n . We denote by $m_t(g)$ the unique point y of g s.t. $d(t, y) = \min\{d(t, x) : x \in g\}$.

Lemma 10. Let $q' \subset q$ be cosets of dimensions j and $(j+1)$. Then $\{t : m_t(q) \in q'\}$ is a hyper-plane.

Definition 13. Let P be a set of hyper-planes. We say that P is *regular* if for any $P' \subset P$:

1. If $|P'| \leq n$ then $\cap P'$ is an $(n - |P'|)$ -dimensional coset.
2. If $|P'| > n$ then $\cap P' = \emptyset$.

Definition 14. Let t be a point and P a regular set of hyper-planes. We say that t *separates* P if for any P' and P'' , distinct subsets⁸ of P of cardinality at most n , $m_t(\cap P') \neq m_t(\cap P'')$.

Lemma 11. Let P be a regular finite set of hyper-planes. Then there is a point t that separates P .

Proof. It suffices to show that there is a point t s.t. $m_t(\cap P') \notin p$ for any $P' \subset P$, $|P'| < n$ and $p \in P - P'$. By Lemma 10 for $q = \cap P'$ and $q' = (\cap P') \cap p$, and since \mathbb{R}^n is not the union of finitely many hyper-planes, there is such a t . \square

Definition 15. For a half-space s , let $s^{(0)}$ denote the half-space adjacent to s , and $s^{(1)}$ denote the half-space s . For a hyper-plane p and a point y not in p , let $h(p, y)$ denote the half-space s s.t. $B(s) = p$ and $y \in s$.

Henceforth, let P be a regular finite set of hyper-planes and t a point that separates P .

Definition 16. For a point y define $P^{(y)} \triangleq \{p \in P : y \in p\}$. Based on P and t we now define, for each point y , a partial function Γ_y from the set of hyper-planes to the set of half-spaces s.t. whenever $\Gamma_y(p)$ is defined then $B(\Gamma_y(p)) = p'$, i.e., $\Gamma_y(p)$ chooses one of the two adjacent half-spaces whose boundary is p . $\Gamma_y(p)$ is defined according to the following two adjacent half-spaces whose boundary is p . $\Gamma_y(p)$ is defined according to the following three cases:

Case 1: $y \notin p$: Define $\Gamma_y(p) \triangleq h(p, y)$.

Case 2: $y \in p$, $p \in P$ and $m_t(\cap P^{(y)}) = y$: let $x = m_t(\cap (P^{(y)} - \{p\}))$. Since t separates P , we have $x \neq y$ and $x \notin p$. Define $\Gamma_y(p) \triangleq (h(p, x))^{(0)}$.

Case 3: Otherwise: $\Gamma_y(p)$ is undefined.

⁸ Note that P' P'' may be empty; if P' (say) is empty, we define $\cap P' \triangleq \mathbb{R}^n$.

Definition 17. Define a *cell* to be any nonempty subset of \mathbb{R}^n which is intersection of finitely many half-spaces. A *cell* of P is a cell of the form $\cap S$ where S is a finite set of half-spaces, $B(S) \subset P$, and S has no adjacent half-spaces.

Lemma 12. Let q be a cell of P with $q = \cap S$ and $B(S) \subset P$, and let $y = m_t(q)$. Then

1. $m_t(\cap P^{(y)}) = y$.
2. $P^{(y)} \subset B(S)$.
3. $s = \Gamma_y(B(s))$ for any $s \in S$.

Proof. Assume $m_t(\cap P^{(y)}) = z \neq y$. Clearly, $d(t, z) < d(t, y)$ and $z \notin q$. Since $\cap P^{(y)}$ is convex, it includes the line interval $[y, z]$. On the one hand, $[y, z] \cap q = [y, y]$. On the other hand, for any $s \in S$: if $y \in B(s)$ then $[y, z] \cap s = [y, z]$, and if $y \notin B(s)$ then $[y, z] \cap s = [y, y'] \neq [y, y]$. Therefore $[y, z] \cap q \neq [y, y]$; a contradiction. This establishes (1).

To establish (2), assume $m_t(\cap(P^{(y)} \cap B(S))) = z \neq y$. All the arguments of the previous paragraph hold and lead to the same contradiction. Hence, $m_t(\cap(P^{(y)} \cap B(S))) = y$. Since t separates P , we have $P^{(y)} \subset B(S)$.

To establish (3), let $s \in S$. Say first that $y \notin B(s)$. By case 1 of Γ_y definition, $\Gamma_y(B(s)) = h(B(s), y)$. This and $y \in s$ imply $\Gamma_y(B(s)) = s$.

Say next that $y \in B(s)$. By case 2, $\Gamma_y(B(s)) = (h(p, x))^{(0)}$ with $x = m_t(\cap(P^{(y)} - \{B(s)\}))$. Consider the line interval $[y, x]$. Since $d(t, x) < d(t, y)$, we have $[y, x] \cap q = [y, y]$. But for any $s' \in S - \{s\}$, $[y, x] \cap s' \neq [y, y]$. Hence, $[y, x] \cap s = [y, y]$; therefore $x \notin s$; this implies $\Gamma_y(B(s)) = s$. \square

Definition 18. We say that $C = \langle S, Z, R \rangle$, a subclass of $(HS^n)^D$, is *regular* if $\mathbb{R}^n, \emptyset \notin S$, S has no adjacent half-spaces, $B(S)$ is regular and⁹ $(\cup B(S)) \cap Z = \emptyset$.

Lemma 13. Let $C = \langle S, Z, R \rangle$ be a finite subclass of $(HS^n)^D$. Then there is a regular subclass of $(HS^n)^D$, $C' = \langle S', Z, R \rangle$, isomorphic to C .

Proof. By induction on $|S|$ and using a Lemma 9. \square

Theorem 3. The class $(HS^n)^D$ admits an unlabeled compression of size n .

Proof. By Lemma 4 it suffices to show that any finite subclass of $(HS^n)^D$ admits such a compression. Let $C = \langle S, Z, R \rangle$ be a finite subclass. By lemma 13 we may assume that C is regular. Set $P \triangleq B(S)$ and let t and Γ be as above.

Let $f : S' \rightarrow \{0, 1\}$ be a sample under C (S' is a subset of S). Let $S'' = \{s^{(f(s))} : s \in S'\}$. Pick $z \in Z$ which is consistent with f . Since $z \in S''$, we have $q \triangleq \cap S'' \neq \emptyset$; i.e., q is a cell of P .

⁹ The last requirement, $(\cup B(S)) \cap Z = \emptyset$, is actually not used. We include this requirement to remark that it is irrelevant whether $(HS^n)^D$ is composed of open or closed half-spaces.

Let $y = m_t(q)$. By lemma 12, Γ_y reconstructs the sample f and y is determined from some S^* , a subset of S' with $|S^*| \leq n$, by $y = m_t(\cap B(S^*))$. \square

By Lemma 1, $(HS^n)^D$ is a \preceq_{emb} -universal class for all Dudley classes whose VC-dim is at most n . This implies:

Theorem 4. *Let C be a Dudley class. Then C admits an unlabeled compression of size $VC\text{-dim}(C)$.*

5. Compression of geometric classes

In this section we consider geometric classes – classes whose relation R is over the real field and is definable there by first-order logic. Geometric classes have been studied by Goldberg and Jerrum [9]¹⁰. They have shown that all geometric classes have a finite VC-dimension and established upper bounds on the VC-dimension of a geometric class as a function of the syntax of its first order formula representation. We show that any geometric class admits an array compression of a finite size. The technique of [9] can be used to establish an upper bound on the size of the array compression. It is yet unknown whether all geometric classes admit a finite-size compression of a stronger type.

Consider the following model:

$$\mathcal{U} = \langle \mathbb{R}, +, \cdot, \geq, r \rangle_{r \in \mathbb{R}}.$$

The language \mathcal{L} of \mathcal{U} has the binary function symbols ‘+’ and ‘ \cdot ’, the binary predicate symbols ‘=’ and ‘ \geq ’ and the constant symbols ‘ r ’ for each $r \in \mathbb{R}$.

Definition 19. Let R be a j -ary relation over \mathbb{R} . R is *definable in \mathcal{U}* if there is a formula $\varphi(v_1, \dots, v_j)$ of \mathcal{L} s.t. for any $r_1, \dots, r_j \in \mathbb{R}$:

$$R(r_1, \dots, r_j) \Leftrightarrow \mathcal{U} \models \varphi(r_1, \dots, r_j).$$

A *geometric class* is a class $C = \langle X, Y, R \rangle$ where $X = \mathbb{R}^l$, $Y = \mathbb{R}^m$ and the $(l+m)$ -ary relation $R'(x_1, \dots, x_l, y_1, \dots, y_m) = R(x, y)$ is definable in \mathcal{U} .

Tarski and Robinson [5] have established that \mathcal{U} admits elimination of quantifiers; that is, for any formula $\varphi(v_1, \dots, v_j)$ of \mathcal{L} there is a quantifier free formula $\hat{\varphi}(v_1, \dots, v_j)$ s.t.

$$\mathcal{U} \models \forall v_1, \dots, v_j : \varphi(v_1, \dots, v_j) \Leftrightarrow \hat{\varphi}(v_1, \dots, v_j).$$

Note that any quantifier free formula of \mathcal{L} can be expressed as

$$f(P_1(v_1, \dots, v_j) \geq 0, P_2(v_1, \dots, v_j) \geq 0, \dots, P_k(v_1, \dots, v_j) \geq 0)$$

¹⁰ Geometric classes are called there “Concept Classes Parameterized by Real Numbers”.

where f is a k -ary Boolean function and each P_i is a polynomial in the variables v_1, \dots, v_j .

Theorem 5. *Let C be a geometric class. Then C admits an array compression scheme of size n for some finite n .*

Proof. Let $c = \langle X, Y, R \rangle$ with $X = \mathbb{R}^l$ and $Y = \mathbb{R}^m$. Let R be expressed by

$$f(P_1(x_1, \dots, x_l, y_1, \dots, y_m) \geq 0, \dots, P_k(x_1, \dots, x_l, y_1, \dots, y_m) \geq 0)$$

as above

For each $1 \leq i \leq k$, the class $C_i \triangleq \langle \mathbb{R}^l, \mathbb{R}^m, P_i(x_1, \dots, x_l, y_1, \dots, y_m) \geq 0 \rangle$ is embeddable in a Dudley class; hence, by Theorem 4, it has an array compression scheme of a finite size, say n_i . By Lemma 7, the class $C' = \langle X', Y', R' \rangle = (C_1 \times \dots \times C_k)^f$ has an array compression scheme of size $n = \sum_{i=1}^k n_i$.

Let $\hat{C} = \langle \hat{X}, \hat{Y}, \hat{R} \rangle$ be the subclass of C' where $\hat{X} = \{ \langle x, x, \dots, x \rangle \in X' \mid x \in \mathbb{R}^l \}$ and $\hat{Y} = \{ \langle y, y, \dots, y \rangle \in Y' \mid y \in \mathbb{R}^m \}$. Clearly, \hat{C} is isomorphic to C . Since \hat{C} has an array compression scheme of size n , so does C . \square

Appendix. $\text{PHS}_0^n \preceq_{\text{gemb}}^* \text{PHS}^{n-1}$

We show here that PHS_0^n , the class of positive half-spaces that pass through the origin, is \preceq_{gemb}^* than PHS^{n-1} , the class of positive half-spaces in \mathbb{R}^{n-1} . To this end, we use the terminology of Section 4 and define PHS_0^n and PHS^n as follows.

Definition A.1. Pick a point y of \mathbb{R}^n different from the origin $\mathbf{0}$. Define:

$$\text{PHS}^n \triangleq \langle \mathbb{R}^n, S^p, \in \rangle; S^p \triangleq \{ s \in H^n \mid \exists x \in \mathbb{R}^n : x \notin s \wedge x + y \in s \}.$$

Define $\text{PHS}_0^n = \langle \mathbb{R}^n, S_0^p, \in \rangle$ as the subclass of PHS^n where

$$S_0^p \triangleq \{ s \in S^p : \mathbf{0} \in B(s) \}.$$

Note that, up to an isomorphism, the classes PHS^n and PHS_0^n are independent of y .

We use the following elementary lemma.

Lemma A.2. *Let l be a line and L a finite set of lines, none of which is parallel to l . Then there is a hyper-plane p such that $p \cap l = \emptyset$ and $p \cap l'$ is a singleton for any $l' \in L$.*

For two distinct points, x and y , let $I(x, y)$ denote the line passing through x and y .

Lemma A.3. $\text{PHS}_n^0 \preceq_{\text{gemb}}^* \text{PHS}^{n-1}$.

Proof. Let $\text{PHS}_0^n = \langle \mathbb{R}^n, S_0^p, \in \rangle$, let y be the selected point according to Definition 20, and let $C' = \langle Z', Q, \in \rangle$ be a finite subclass of PHS_0^n .

Since $l(\mathbf{0}, y) \cap B(s) = \{\mathbf{0}\}$ for any $s \in Q$, there is a finite subclass of PHS_0^n , $C = \langle Z, Q, \in \rangle$, isomorphic to C' s.t. $Z \cap l(\mathbf{0}, y) = \emptyset$. By Lemma 14, there is hyper-plane p s.t. $p \cap l(\mathbf{0}, y) = \emptyset$ and $p \cap l(\mathbf{0}, z)$ is a singleton (say $\{\hat{z}\}$) for any $z \in Z$.

Consider the class $D \triangleq \langle p, G, \in \rangle$ where $G = \{p \cap s : s \in S_0^p\}$. Clearly, D is isomorphic to PHS^{n-1} .

To establish a generalised embedding of C into D , define the functions $\pi : Z \mapsto p$, $\tau : Q \mapsto G$ and $\sigma : X \mapsto \{0, 1\}$ as follows. $\pi(z) \triangleq \hat{z}$, $\tau(s) \triangleq s \cap p$, and $\sigma(z) = 1$ iff $\mathbf{0}$ is in the line-interval $[z, \hat{z}]$.

Since a half-space and its complement are convex, for any $z \in Z$ and $s \in Q : z \in s$ iff $[\pi(z) \in \tau(s) \Leftrightarrow \sigma(z) = 0]$. That is, (π, τ, σ) constitute a generalised embedding of C into D . \square

References

- [1] N. Alon, P. Frankl, V. Rodel, Geometrical relation of set systems and probabilistic communication complexity, Proc. Symp. on the Theory of Computing, 1985, pp. 277–280.
- [2] S. Ben-David, N. Bshouty, E. Kushilevitz, A composition theorem for learning algorithms with applications to geometric concept classes, ECCC Technical Report TR96-059 (1996), and Proc. 29th annual ACM symp. on theory of computing, (STOC'97), pp. 324–333.
- [3] S. Ben-David, N. Eiron, H.U. Simon, On the embeddability of finite VC classes in classes of halfspaces, Technion CS department manuscript, 1997.
- [4] A. Blumer, A. Ehrenfeucht, D. Haussler, M. Warmuth, Learnability and the Vapnik-Chervonenkis dimension, J. Assoc. Comput. Mach. 36(4), (1989) 929–965.
- [5] C.C. Chang, H.J. Keisler, Model Theory, North-Holland, Amsterdam, 1990.
- [6] R.M. Dudley, A Course on Empirical Processes, Lecture Notes in Mathematics, vol. 1997, Springer, New York, 1984.
- [7] S. Floyd, On space bounded learning and the Vapnik-Chervonenkis dimension, Ph.D. thesis, International Computer Science Institute, Technical Report, TR-89-061.
- [8] S. Floyd, M. Warmuth, Sample compression, learnability, and the Vapnik-Chervonenkis dimension, Mach. Learning, 21, (1995) 1–36.
- [9] P.W. Goldberg, M. Jerrum, Bounding the Vapnik-Chervonenkis dimension of concept classes parametrized by real numbers, Mach. Learning, 18 (1995) 131–148.
- [10] M.C. Laskowski, Vapnik-Chervonenkis classes of definable sets, J. London Math. Soc. 45(2), (1992) 377–384.
- [11] N. Littlestone, Mistake bounds and logarithmic linear-threshold learning algorithms, Ph.D. thesis, U.C. Santa Cruz, March 1989.
- [12] N. Littlestone, M. Warmuth, Relating data compression and learnability, Unpublished manuscript, (1986).
- [13] B.K. Natarajan, On learning sets and function, Mach. Learning 4 (1989) 67–97.
- [14] L. Pitt, M. Warmuth, Prediction preserving reducibility, J. Comput. System Sci. 41 (1990) 430–467.
- [15] N. Sauer, On the density of families of sets, J. Combin. Theory. Ser. A 13 (1972) 145–147.